

Building a Pragmatic Semantic Web

Harith Alani, Wendy Hall, Kieron O'Hara, Nigel Shadbolt, and Martin Szomszor,
University of Southampton

Peter Chandler, CA Technology

A practical approach to adopting Semantic Web technologies enables large organizations to share data while achieving clear private as well as public reuse benefits.

Many real-world tasks require the acquisition and integration of information from a distributed set of heterogeneous sources. Hence, there's no shortage of opportunities for applications using Semantic Web (SW) technologies. The power of publishing and linking data in a way that machines can automatically interpret

through ontologies is beginning to materialize.¹ However, market penetration level is relatively low, and it's still no routine matter for an enterprise, organization, governmental agency, or business with large distributed databases to add them to the Web of linked and semantically enriched data. It's also probably fair to say that many organizations still view the SW with some skepticism. In part, they may suspect that they're expected to pioneer an approach in which quick wins are few. Moreover, cost and privacy issues arise when ever-increasing amounts of information are linked into the Web.

Perhaps understandably, most academic work has focused on the global public gains of adopting SW technologies. Equally understandable, many organizations are wary of being early adopters if public gains are all they can anticipate. If that were the case, the SW would more likely be considered a corporate "social responsibility" than a business-enhancing advance in information management. We argue, however, that the SW offers local, private gains for the individuals and organizations that link their data and information. These gains are inseparable from the global, public gains, and cost-benefit analyses must consider both types to be accurate.

We discuss an approach to the initial stages of building SW applications. We designed the approach to be practical, cost effective, fast, and ap-

pealing to organizations that can't afford to neglect the bottom line, take many risks with their information, or think only in the long term. Such organizations must often negotiate a variety of obstacles—including scalability, different terminologies, and diverse formats—to efficiently share and reuse their information. We present our experiences with one project that targets the public sector (AKTivePSI) and one in the private sector (MRO Expressway).

Attracting organizations to the SW

Introducing any new technology to an organization requires careful management. However, introducing SW technology also involves selling the SW to an organization that can easily see the costs of conversion seem overwhelming while the benefits are less clear.²

We met with several organizations and listened to their views and concerns about SW costs and benefits. Table 1 presents some misconceptions about the demands on users. To render this technology less daunting to potential users, we developed the following four principles.

Minimize disruption to existing infrastructure

Making a complete, fast transition to semantic knowledge bases (KBs) is unnecessary and impractical in the short to medium term. Organizations

Table 1. Common misconceptions about the Semantic Web.

Misconception	Reality
Everyone must agree on the same terminology to enable data and information sharing.	Different groups use different terminologies, with lightweight mappings where required to ease sharing and communication.
Ontologies are typically large and complex.	Heavyweight and complex ontologies encode domain knowledge. Applications don't always require such ontologies; their data is often well represented using relatively lightweight ontologies.
Ontologies are expensive to design, build, and maintain.	Some ontologies encode a great deal of domain knowledge and can be expensive to build. In these heavyweight ontologies, the larger the potential user community the more it offsets the cost of construction. Lightweight ontologies can have wide applicability and can be cost effective to build in terms of overall utility to the community. ³
Information and data are taken out of current knowledge management practices, expensively converted to RDF, and replaced with new standards and technology.	RDF creation can be automated, using simple scripts, APIs, or conversion languages such as GRDDL (Gleaning Resource Descriptions from Dialects of Languages). Data and information can be kept in their current formats, and cached or exported in RDF.
Providing access to data and information benefits consumers and competitors but offers no quick wins for the provider.	In the long run, exposing data and information will provide gains for owners as well as the whole network, just as exposing documents provided gains when the original Web took off. In the short term, facilitating information reuse generates quick wins for organizations with a large quantity of distributed legacy data in heterogeneous formats.
The promiscuous release of data and information will be a privacy nightmare.	Standards are being developed to control access and reuse policies. In the meantime, as with conventional databases and Web technologies, organizations can pick and choose what data and information to expose and share.

need to know that adopting SW technology doesn't mean throwing away existing database and information technologies. Reengineering information processing is a complex task facilitated by special conditions and structures that vary across countries and cultures. So, the first principle must be to minimize disruption to participants' existing data and information flows and models.

An important part of our task is to show that SW adoption costs are relatively low. We've demonstrated how to use simple scripts to convert data into RDF triples with enriched interpretations, and adopted the approach of caching the given databases into triple stores. An alternative is to use a technology such as D2RQ (<http://sourceforge.net/projects/d2rq-map>), which enables layering an ontology on top of a non-RDF database, thus removing the need to cache or change the original database's structure.

We've found simulating a real-life scenario useful. It lets us show that what we build and do in our lab environment can be done the same way outside it. For example, we treat the KBs as if the participants hosted them. This also helps bridge the credibility gap: it's hard to imagine the SW's utility in advance of its global availability. The document Web faced similar challenges before the World Wide Web came into its own. Sector-specific SWs and SW intranets can be compelling illustrations.

Use small, well-focused ontologies

It's not realistic to assume that an organization will build one monolithic ontology for all its data and information or that different organizations will agree on one semantic model. Constructing a new ontology for each information asset and designing it to represent only what's stored in a particular database has proved a good intermediate step. We can then map these small ontologies to form a small SW.

Ontologies vary according to their degree of formalization, their purpose, and the subject matter they represent. One recommended first step toward building an ontology is to scope its domain to make sure it doesn't grow larger than necessary. The appropriate size depends on the ontology's purpose and domain. Some, such as the Gene Ontology (GO, www.geneontology.org), are designed to represent entire domains and tend to be very large. Other ontologies might serve the needs of specific applications and can be smaller. Still others are data dependent and built mainly to represent and improve accessibility to a data collection. The smaller and simpler the ontology, the less expensive and time-consuming it is to develop and maintain.

Constructing ontologies requires skills such as modeling knowledge and expertise. Small organizations or organizations

on limited budgets, such as government bodies, worry about the possible high cost of building complex knowledge structures. We've been able to demonstrate how much organizations can achieve with practical ontologies that are scaled down to fit individual information assets rather than entire domains. They can gradually link such ontologies together to facilitate data sharing. They might later require more elaborated ontologies to further automate ontology mapping or to check for data inconsistency. However, starting small is important to massage perceptions of affordability for most organizations.

Show added value

Providing better access to information isn't enough to completely win information providers' interest, support, and active participation. You must also show examples of where and what the added value of integration and shared access will be. Most organizations have needs—and sometimes laborious procedures—for acquiring data and information from other sources. A semantically enabled content-exchange channel offers direct benefits with respect to consistency checking, relative ease of integration and distributed querying, and efficient data and information exchange and merging.

Integration from multiple content sources adds to the value of knowledge augmenta-

tion and verification. The integration offers useful insights into data set quality for the provider involved, helping to uncover errors and inconsistencies and highlighting knowledge gaps.

Preserve provenance and privacy

Many agencies and institutions are instinctively secretive about their data. The SW vision is to remove human processing from knowledge acquisition as far as is feasible. However, the idea of publishing data without controlling its presentation context is very new in most industrial and government circles (although an ancient problem dating back at least to Plato's *Phaedrus*). These agencies need assurance that SW technology will let them choose what to share and what to keep private.

Some of the organizations we met with expressed great concern about possible misuse of data or information once the SW technology enabled access and reuse. To ameliorate these concerns, we transfer each resource we received into a separate KB with its own ontology. This approach eliminates any risks of contamination from one database to another. Furthermore, each ontology contains a few classes and properties to represent the data source, such as the supplier's name, data set name, and date supplied. We also attach source information to all triples in the triple store.

Privacy is a complex issue. Many of us are prepared to surrender our privacy for gains in efficiency or monetary benefit; others defend personal privacy as a pillar of a liberal democratic society. Unless and until such political dilemmas are resolved, organizations must carefully consider how far to exploit SW and other information technologies.⁴

The World Wide Web Consortium (W3C) is developing technologies and protocols to create a policy-aware Web. The W3C standards that eventually emerge from this process will enable information users, owners, and subjects to express policies for information use and negotiate about them.⁵

Constructing SW applications

We followed these principles in building two SW applications, one in the public and the other in the private domain. Once we had the application data for these projects, the building process involved similar steps in both domains. We summarize them here.

Construct ontologies

To ensure low complexity in the ontologies we built for the provided data sets, we limited their scope and size. Small ontologies are cheaper and easier to build, maintain, understand, use, and commit to. None of the participating organizations' databases required a large number of concepts and relationships to represent the stored information. We were able to show that ontologies aren't hard to build if their purpose is representing databases and information assets of circumscribed scope. We also showed that they don't require consensus on a common vocabulary. With ontology-mapping techniques, local terminologies prove sufficiently useful.

Integration from multiple content sources offers useful insights into data set quality, helping uncover errors and highlighting knowledge gaps.

The average number of classes in our ontologies was 30, with a median of 10 classes.

Generate RDF

From an ontology, we created instances by running simple scripts over the data to produce RDF. Initially, we generated the scripts manually for a particular database/ontology pair. When possible, we reused these scripts across similar databases and ontologies. Although we built the scripts manually, a framework for semiautomatic script generation is conceivable. We used the Jena API (<http://jena.sourceforge.net>) to write most of the scripts, which made them reusable and easy to tune for new data sets and ontologies. This process demonstrated the relative ease of converting legacy data to RDF using simple and free SW technology.

Although we needed small ontologies to interpret the data, we also needed a scalable KB to hold the millions of RDF triples generated. To store the generated RDF files, we

used 3Store (www.aktors.org/technologies/3store), an RDF triple store developed in the Advanced Knowledge Technologies (AKT) project (www.aktors.org). This triple store provides a SPARQL (SPARQL protocol and RDF query language) endpoint—that is, a servlet that accepts SPARQL queries and returns XML results.

By publishing RDF in accordance with best practices,⁶ this data becomes viewable with general-purpose RDF browsers such as Tabulator (www.w3.org/2005/ajar/tab).

Uniform Resource Identifiers (URIs) play a fundamental role in SW publishing. All SW entities of interest, such as information resources, real-world objects, and vocabulary terms, need a URI reference. Once we have URI references, we can insist that they should be dereferenceable.⁷ This means a person or an application can look up a URI over the Web and retrieve information about the identified resource.

Migrate to the Semantic Web

URI reuse increases connectivity between published data, facilitating discovery of related data.⁷ However, sometimes it's unclear who should reuse whose URIs, especially when organizations aren't experienced in this field and are unaware of other efforts to enrich data semantically. Nevertheless, you can connect ontologies to each other by mapping their equivalent URIs.

Ontologies facilitate integration by using soft mappings between concepts and instances that queries or data browsers can follow to find similar or duplicated entities. We've used the special `owl:sameAs` property to link any mapped entities. Connecting our KBs in this way let us provide much greater flexibility and querying power than the original data structures allowed.

Because one key aim of this research is to show the added value of using SW technology for publishing and using data, we had to show how to form a bigger semantic network by integrating the KBs containing all the project data. Accordingly, we performed mappings of both local ontologies and their instances.

Even though automatic ontology mapping has been a research focus for many years and many tools are available for it, our ontologies' relatively small size made it easier to map them manually than to correct automated mappings. The mapping process wasn't difficult, although the participating organizations' domain expertise provided

important input to it. As we will show later, mapping doesn't have to be complete to be useful. You can draw significant value from mapping even a small number of concepts.

Our data-centric SW approach makes mapping the instance data to each other useful as well. These mappings must be automated because there are usually many instances to map. We do this with simple scripts that search for duplicates of specific instance types (for example, postcodes and airplane models). An `owl:sameAs` link can be added automatically between the corresponding instances once we or the automated tool finds such a mapping. These processes create several files that contain RDF `owl:sameAs` triples linking various parts of the data. We store these files separately from the data and invoke them in queries. To retrieve data from the KB, our applications use SPARQL queries. Because the ontologies and data have been linked as described, it is possible to extract information from multiple data sources.

AKTivePSI

The UK Office of Public Sector Information (OPSI) manages all the government's intellectual property, including setting standards, delivering access, and encouraging the reuse of public-sector information. OPSI also regulates holders of public-sector information (PSI), such as the Met Office and the Ordnance Survey, in their information-trading activities.

Information policy has developed rapidly in the UK over the past five years, with Freedom of Information legislation as well as an EU directive on opening access to PSI, but no large-scale research has addressed the potential for reuse with SW technologies and approaches. OPSI initiated a small project, AKTivePSI,⁸ to show what could be achieved if public-sector information was made available for reuse in an enabling way.

Throughout the project, we consulted regularly with many governmental organizations, including the London boroughs of Camden (www.camden.gov.uk) and Lewisham (www.lewisham.gov.uk), Ordnance Survey (www.ordnancesurvey.co.uk), Stationary Office (TSO, www.tso.co.uk), Met Office (www.metoffice.gov.uk), Environment Agency (www.environment-agency.gov.uk), and Office of National Statistics (ONS, www.statistics.gov.uk).

Direct outcomes of AKTivePSI include the following:

- The *London Gazette* ([\[online.co.uk/home.spx?geotype=London\]\(http://online.co.uk/home.spx?geotype=London\)\) is building OWL ontologies to represent parts of ITS data and is working toward publishing this data in RDF.](http://www.gazettes-

</div>
<div data-bbox=)

- The OPSI oversaw the development of a URI schema, which it's using to generate URIs for government legislation and copyright statements.
- The Camden Borough Council added a SW engineer to its staff force to help the council to join the SW.
- The Ordnance Survey is continuing its SW work and research; it has already built several ontologies and released several data sets.

Initially, the project aimed to draw to-

A small project, AKTivePSI, showed what could be achieved if public-sector information was made available for reuse in an enabling way.

gether a sufficiently large set of heterogeneous information from a selection of public-sector organizations to

- explore how SW technology could help turn government information into reusable knowledge to support e-government;
- investigate the best practical approaches to achieve this goal, in terms of collecting data and constructing ontologies;
- show how data can be integrated and identify existing government taxonomies that are useful for this task; and
- provide evidence of the added value from undergoing his process.

To help focus the requests for content, we collected information from the geographical area covered by the two participating London boroughs.

Public-sector data sets

Several participating organizations made databases available for the project (table 2).

We developed scripts to convert this data to RDF automatically, in correspondence with the designated organizations' ontologies. In total, we constructed 13 ontologies, one for each data set in table 2.

Mappings

For example, we developed two ontologies for data sets from the Lewisham Borough Council. Each ontology has classes representing property, address, and postcode. We linked these concepts with `owl:sameAs` to indicate that they represent the same concepts.

Many simple mappings were also available, such as mapping the concept Premises from the Camden's Food Premises ontology to the Property class in its Land and Property ontology. Although simple, such mappings can still be powerful. The postcode instance N6 6DS in one KB mapped to the instance pc_N66DS in another. Because these instances really did refer to the same object, we could infer much more information about it by noting the identity. In fact, we found that simply linking to one data object (the postcode) was enough to glean useful information from various data sets for the creation of interesting mashup applications.

Mashing up distributed knowledge bases

Once data is available in easily parsable and understandable formats such as RDF, mashups become much easier to generate by searching RDF KBs and mashing up data on the fly—a clear benefit of linking. We created two such mashups in AKTivePSI to demonstrate these benefits and the relative ease of constructing them from semantically represented knowledge.

The Camden Food Premises data set gives information about hygiene inspections and health risks of various premises in the Camden area that handle food. The risk categories range from A, which is high risk, to E, which is low risk. The category is based on the premises' cleanliness, compliance with regulations, type of preparation that's performed, and so on. The Food Premises database contains much information on these properties, but displaying the information on a map is difficult because the data set lacks geographical coordinates.

However, the Ordnance Survey's Address Layer and Points of Interest (PointX) data sets do contain geographical coordinates for businesses and properties. The instance

Table 2. Data sets provided to AKTivePSI, the number of RDF triples we generated for each data set, formats, and a description of what the data is about.

Data set	No. of RDF triples	Format	Data description
Camden Borough Council			
Land and Property Gazetteer	2.3M	Excel	Properties in Camden, full address, coordinates, and type (residential/nonresidential/mixed)
Food Premises	84K	Excel	Food-related premises in Camden, their business names, hygiene inspection results, addresses, and classifications (for example, restaurant, school, bar)
Local Businesses	170K	Excel	Businesses in Camden, names, addresses, contact info, and business type
Licenses	100K	MSSQL	Licenses for businesses in Camden, their addresses, license types, and expiration dates
Councillors and Committees	29K	Excel	Councillors and committees, subcommittees, who sits on which committee, and councillors' personal information
Meeting minutes	106K	Text	Web pages of committee's meeting minutes
Lewisham Borough Council			
Land and Property Gazetteer	4M	Excel	Properties in Lewisham, their full addresses, and coordinates
Property Tax Bands	10K	Excel	Tax property references, description, rate payers, rate value, and single-string addresses
Ordnance Survey (data for Camden and Lewisham only)			
Address Layer 1	768K	XML	Data about buildings, addresses, and coordinates
Address Layer 2	11.7M	XML	Data about buildings, addresses, and coordinates, and building classifications (for example, hospital, university)
PointX (Points of Interest)	467K	XML	Various landmarks and businesses, with names, addresses, and coordinates
The Stationery Office <i>London Gazette</i> (entire database was provided, but only that below was used)			
Administration Notices	120K	Text	Notices for the appointment of administrators for corporate insolvencies
Deceased Estates	3.2M	Text	Death notices of individuals, names, addresses, description and date of death, and address of representatives

mapping of postcodes we performed earlier helped reduce our search space for finding matching addresses in the data sets. Indeed, once we found matches, we could assert that they were the same, thereby avoiding the need to search again.

To create the mashup, we wrote several SPARQL queries that searched for each premise's address from the Food Premises data set in each of the two Ordnance Survey data sets. When we found a match, we retrieved the coordinates and displayed the premise on a Google map. The information from Food Premises together with the mapping between the data sets provided extra context to instances from either data set. The PointX data set gains access to the food premises' risk level (as well as the implicit knowledge

that the premises are used for preparing food), and the Food Premises data set garners exact coordinates for the premises. Figure 1 shows a simple Google Maps mashup that uses the mapping to provide a visual display of the Food Premises data set.

This type of mashup promotes public awareness and commercial competition. For example, one particular business that was placed within the high-risk category has glowing customer reviews on restaurant review sites across the Internet.

Inconsistencies

Data and information integration from multiple sources adds the value of knowledge augmentation and verification. Integrating data sets can give the data provider useful in-

sights into a data set's quality. For example, the Ordnance Survey's Address Layer 2 data set provides a list of businesses, including their addresses and their geolocations, and the PointX data set provides similar information. However, we found that the two lists of businesses didn't match. For instance, some businesses were in one data set but not the other. In some cases, the PointX data set contained several businesses listed at the same address, while the Ordnance Survey Address Layer 2 listed only one. Was this an error? The data set lacked temporal information, so perhaps it held both former and the current tenants. Or perhaps several businesses occupy different floors in the same building. Inferring an answer is difficult, but at least the integration can



Figure 1. Google Maps mashup of the Camden Food Premises data set. The mashup results from mapping the Food Premises data to the Ordnance Survey's Address Layer II and PointX data sets.

flag possible quality issues for information managers to resolve.

MRO Expressway

The airline industry regularly gathers and publishes data about aircraft orders, sales, registrations, engine specifications, compatibility, repair shop locations, and so on. However, the data comes in various textual forms with little machine-readable structure. We developed an application to store and manipulate a wide range of such data, such as historical aircraft deliveries, new aircraft sales, engine types,

MRO (Maintenance, Repair and Overhaul) business details, thermal spray coatings, and market details. Users can query the consolidated data to answer questions such as the number and type of engines in MRO shops in any geographical area.

MRO Expressway uses a few information sets and a simple model to forecast the worldwide civil-engine repair business.⁹ We demonstrated this capability in great detail for one particular activity—that is, thermal spray coatings. The numbers generated by our application agree well with other published forecasts (for example, see Aero-

Strategy Management Consulting at www.aerostrategy.com). However, the technology could also be a platform for many other application domains, such as design.

Strategically, we intended MRO Expressway to show what a linked SW could offer a particular industrial sector. Such sector-based proofs of concept must themselves be capable of being built cost-effectively and according to the principles we've described.

System objectives

The current drive to minimize emissions has increased attention to component coatings. Closely defining coatings and their effects on performance is a difficult task that would benefit enormously from a system containing coating types and performance data. Designers and engineers could examine materials, coatings, and service options to identify specific gaps in knowledge. The system could also provide a framework for building a strategically important design tool.

MRO Expressway has three principle objectives:

- Consolidate data from multiple heterogeneous sources into a single representation that facilitates extracting information beyond what the individual sources can provide.
- Provide a graphical interface to view and explore data.
- Forecast future MRO recoating business by estimating the number of planes and engines in operation at a regional level.

These objectives are typical of the gains

Table 3. Data sets provided to the airline industry Maintenance, Repair, and Overhaul application.

Data set	No. of RDF triples	Format	Data description
Airbus			
Airbus-Orders	29K	Excel	Airlines, aircraft, and number ordered, delivered, and operational
Boeing			
Boeing-Deliveries	140K	Excel	Airline, country, region, model, engine fitted, order date, and number ordered
Engine Yearbook 2007			
Aircraft-Engines	768	Excel	Aircraft, model, type, number of engines, and compatible engines
Engine-Overhaul	1K	PDF	Company name, address, and auxiliary-power-unit types
CIA			
Factbook	205K	RDF	Country coordinates, flags, demographics, and national statistics on communications and economics

that users can expect from linking and semantically enriching data.

Aviation industry data sets

Information describing airplane models, engine types, and the MRO shops that can repair them is available from several public sources. Manufacturers such as Airbus and Boeing provide detailed information on the planes each airline orders in a variety of formats, such as Excel spreadsheets and PDF documents. We developed several scripts to convert this information to RDF. Table 3 outlines the data sets used in this application, including the number of triples created, the source data format, and a summary of the information they contain. As with AKTivePSI, the overhead for building and mapping the ontologies, and writing the scripts to generate instance data wasn't excessive.

In the first stage, we created a suitable ontology to capture the semantics and structure of each data source. Although much of the data is replicated in different sources, in terms of instances (for example, that American Airlines is from the US) and concepts (for example, aircraft, engine), the data granularity differed significantly. We therefore built different ontologies to better suit each data set.

For example, the Boeing-Deliveries data set specifies the date an order was made, the airline that made it, the airplane model and quantity ordered, and the engine. But the Airbus-Orders data set provides only summaries for each airline, stating the number ordered of each model, the numbered delivered, and the number currently operational. The decision to build different ontologies meant some overlap and redundancy, but it also simplified task planning strategically and pragmatically. The benefits—unlike the costs—increased.

After converting the source data into RDF, we used the owl:sameAs property to link concept instances from each data set that refer to the same entity. This stage provides the power to query over multiple data sources simultaneously. For example, by linking instances of airlines and countries between the Boeing and Airbus data sets, users can query the knowledge base for all orders made by region, country, or airline.

Data presentation

As with AKTivePSI, MRO Expressway centers on a straightforward Google map in-



Figure 2. The Maintenance, Repair, and Overhaul (MRO) Expressway GUI. The Google map provides a straightforward interface to the information available through data in the RDF triple store.

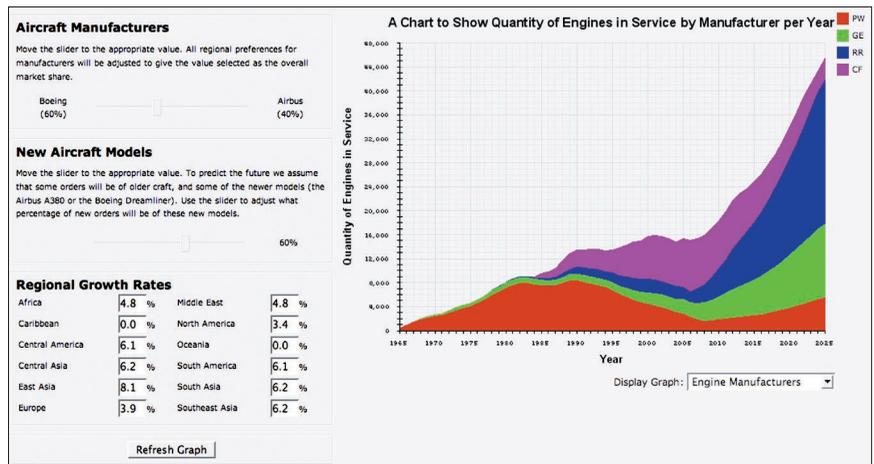


Figure 3. The MRO Expressway GUI. A graph of future business is based on industry parameters, with user-set values on the left side.

terface (see Figure 2). The map area depicts various pieces of information held in, or calculated from, data stored in the triple store. In Figure 2, the map area shows the locations of MRO shops, which users can select to view additional information, such as the company name and the engine models it can repair. In the bottom part of the screen, a tabular data browser lets users inspect the data. The interface contains several methods for presenting data, with views to highlight regional information, repair shops, or quantities of MRO recoating business generated by region and time period (including levels of future business based on a forecasting algorithm, as shown in Figure 3).

We built both the AKTivePSI and MRO Expressway applications to cover information stores of real-world complexity and heterogeneity. This process involved selling the use of SW technology to clients who were, if not skeptical, at least understandably risk averse. Persuading organizations to be in the SW vanguard can be difficult, especially if your strategy is to enlist them in a public service. To convince them to devote scarce resources to the task, you must highlight the private gains. Because the main SW source papers emphasize the public benefits over the private, information officers could be forgiven for thinking they face a trade-off between



**RUNNING IN CIRCLES
LOOKING FOR
A GREAT
COMPUTER JOB
OR HIRE?**

The IEEE Computer Society Career Center is the best niche employment source for computer science and engineering jobs, with hundreds of jobs viewed by thousands of the finest scientists each month in *Computer* magazine AND online!

- > Software Engineer
- > Member of Technical Staff
- > Computer Scientist
- > Dean/Professor/Instructor
- > Postdoctoral Researcher
- > Design Engineer
- > Consultant
- > Hardware Engineer

IEEE
computer
society

<http://careers.computer.org>

The IEEE Computer Society Career Center is part of the *Physics Today* Career Network, a niche job board network for the physical sciences and engineering disciplines. Jobs and resumes are shared with four partner job boards - *Physics Today* Jobs and the American Association of Physics Teachers (AAPT), American Physical Society (APS), and AVS: Science and Technology of Materials, Interfaces, and Processing Career Centers.

private expenditure and public benefit.

These applications showed clear benefits at relatively low costs. Building small, data-centric ontologies was an easily achievable goal. We think the lessons will interest the wider SW field, where arguments continue regarding the cost of developing and maintaining ontologies, and will contribute to Web science as it investigates the complex interactions between the Web and the offline world.¹⁰ The technologies we've discussed are precursors to the next level of machine information processing, as we move from a

The Authors

Harith Alani is a senior research fellow at the School of Electronics and Computer Science, University of Southampton. His research interests include the Semantic Web, ontologies, and Web 2.0. Alani received his PhD in computer studies from the University of Glamorgan. He is a fellow of the Web Science Research Initiative. Contact him at h.alani@ecs.soton.ac.uk.

Peter Chandler is managing director of the CA Technology consultancy company (www.surfaceweb.com). His interests are in surface engineering, nanotechnologies, advanced materials, and processes in the aeronautics, energy, and biomedical sectors. Chandler received his PhD in surface physics from the University of Southampton. He's a chartered engineer, a chartered scientist, and a member of the Institute of Materials and the Institute of Directors. Contact him at peterchandler@btinternet.com.

Wendy Hall is a professor of computing in the School of Electronics and Computer Science, University of Southampton. Her research interests are Web science, the Semantic Web, hypermedia, and multimedia. Hall received her PhD in pure mathematics from the University of Southampton. She is a vice president of the ACM, a senior vice president of the Royal Academy of Engineering, fellow of the British Computer Society, a fellow of the IEEE, and the director of the Web Science Research Initiative. Contact her at wh@ecs.soton.ac.uk.

Kieron O'Hara is a senior research fellow at the School of Electronics and Computer Science, University of Southampton. His interests are in the politics, sociology, and epistemology of technology, especially the World Wide Web and the Semantic Web. O'Hara received his D.Phil. in the philosophy of artificial intelligence and psychological explanation from the University of Oxford. He is a fellow of the Web Science Research Initiative and a trustee of the Agora think tank. Contact him at kmo@ecs.soton.ac.uk.

Nigel Shadbolt is a professor of artificial intelligence in the School of Electronics and Computer Science, University of Southampton. His research interests are the Semantic Web, Web science, and knowledge technologies. Shadbolt received his PhD in artificial intelligence from the University of Edinburgh. He is a fellow of the British Computer Society, the director of the Web Science Research Initiative, and the chief technology officer for Garlik Limited. Contact him at nrs@ecs.soton.ac.uk.

Martin Szomszor is a research fellow at the School of Electronics and Computer Science, University of Southampton. His research interests are the Semantic Web, Web 2.0, and complex systems. Szomszor received his PhD in computer science from the University of Southampton. Contact him at mns2@ecs.soton.ac.uk.

Web linked primarily through documents to a Web linked at a much finer granularity of content. ■

References

1. N. Shadbolt, T. Berners-Lee, and W. Hall, "The Semantic Web Revisited," *IEEE Intelligent Systems*, vol. 21, no. 3, 2006, pp. 96–101.
2. H. Alani et al., "Towards a Killer App for the Semantic Web," *Proc. 4th Int'l Semantic Web Conference (ISWC 05)*, LNCS 3729, Springer, 2005, pp. 829–843.
3. T. Berners-Lee, "The Fractal Nature of the Web," World Wide Web Consortium, 1998–2006, www.w3.org/DesignIssues/Fractal.html.
4. K. O'Hara and N. Shadbolt, *The Spy in the Coffee Machine: The End of Privacy as We Know It*, Oneworld, 2008.
5. D. Weitzner et al., "Creating a Policy-Aware Web: Discretionary, Rule-Based Access for the World Wide Web," *Web and Information Security*, E. Ferrari and B. Thuraisingham, eds., Idea Group, 2005, pp. 1–31.
6. C. Bizer, R. Cyganiak, and T. Heath, "How to Publish Linked Data on the Web," 2007, <http://sites.wiwiw.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial>.
7. T. Berners-Lee, "Linked Data," World Wide Web Consortium, 2007, www.w3.org/DesignIssues/LinkedData.html.
8. H. Alani et al., "Unlocking the Potential of Public Sector Information with Semantic Web Technology," *Proc. 6th Int'l Semantic Web Conf. (ISWC)*, LNCS 4825, Springer, 2007, pp. 708–721.
9. P. Chandler et al., "Advanced Knowledge System for Coatings and the Gas Turbine MRO Industry," to be published in *Proc. Int'l Thermal Spray Conf. and Exposition*, 2008, http://eprints.ecs.soton.ac.uk/15460/1/itsc_final2.pdf.
10. T. Berners-Lee et al., "Creating a Science of the Web," *Science*, vol. 313, no. 5788, 2006, pp. 769–771.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.